Molecular Characterization and Homology Modeling of Intercellular Adhesion Regulatory (IcaR) Proteins in Biofilm-Producing *Staphylococcus* Species

Donmig Jaula Cunanan, Patricia Cristina Noora Gonzalvo, Matthew Lauren Cumigad Gochuico, Charlotte Diane Villareal Falzado, Acelah Estrada Cruz, Emmanuel Pastolero Dejapa, Chloe Reyna Mae Ybañez Godin, Trisha Nicole Reyes Fenix, Mary Rose Ferrer Lirio

Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University-Manila, Nicanor Reyes St., Sampaloc, Manila, Metro Manila, PHILIPPINES.

ABSTRACT

Background: Despite the threat biofilms pose to inpatient care, there is a lack of investigations towards the IcaR protein that regulates the intercellular adhesion locus, which contributes to biofilm formation via polysaccharide intercellular adhesin synthesis. **Aim:** Since only IcaR proteins of *S. aureus* and *S. epidermidis* have been experimentally determined, the study aimed to produce theoretical yet realistic models of other relevant *Staphylococcus* species (*S. xylosus, S. haemolyticus, S. argenteus, S. saprophyticus, S. caprae,* and *S. capitis*) to further understand their functional capabilities in biofilm formation. **Materials and Methods:** IcaR protein sequences retrieved from NCBI GenBank were subjected to physico-chemical profiling using ProtParam (ExPASy), secondary structure determination through SOPMA, and homology modeling using SWISS-MODEL, i-TASSER, and Phyre2. The best models determined via Ramachandran Plot analysis were refined through the GalaxyWEB server. The models were visualized and analyzed with RCSB Molstar. **Results:** The IcaR proteins of *S. aureus, S. saprophyticus,* and *S. argenteus* displayed similar biofilm regulatory functions. In contrast, *S. epidermidis, S. caprae, S. haemolyticus, S. xylosus,* and *S. capitis* are exceptionally homologous, suggesting similar *ica* Operon regulations. The conserved structures were proportional to their lineage proximity in the Phylogenetic tree. The N-terminal had retained most of its homology while the C-terminal region has produced the greatest variations between the selected *Staphylococcus* species, indicating variation in dimer stability and ligand susceptibility. **Conclusion:** The theoretical data generated by this study enhances the existing literature on biofilm regulation. Such knowledge can be applied to develop interventions that target biofilm-related nosocomial infections.

Keywords: Biofilm, Homology Modeling, IcaR Protein, Molecular Characterization, Staphylococci.

Correspondence:
Ms. Mary Rose Ferrer Lirio,
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University – Manila, Nicanor Reyes St., Sampaloc, Manila, Metro Manila, PHILIPPINES.

Email: mlirio@feu.edu.ph

INTRODUCTION

Biofilms are complex and multicellular communities of micro-organisms that adhere to surfaces and are encased in a matrix of Extracellular Polymeric Substances (EPS). This matrix immobilizes the "biofilm cells" to clump together, enabling communication and cellular processes.[1] Biofilms are commonly found in natural and man-made environments with the ability to form on a variety of surfaces, including medical devices.[2] The National Institute of Health in 2020 reported that 65% of microbial infections, including prosthesis and implantable device-related infections, are related to biofilm formation.[3] For this, biofilms play a pivotal role in causing Hospital-Acquired Infections

SCAN QR CODE TO VIEW ONLINE

www.ajbls.com
DOI: 10.5530/ajbls.2023.12.41

(HAIs) that are difficult to eradicate and treat with conventional antimicrobial therapy, leading to increased hospitalization, higher healthcare costs, and even death.[4] Recent studies have identified Gram-positive Staphylococci as the leading cause of Device-Related Infections (DRIs) in hospitals. Among the genus, *S. aureus* and *S. epidermidis* are known for causing persistent infections and resisting treatment by creating biofilms on indwelling medical devices.[5] With a significant rate of mortality, infections of artificial heart valves produced by both *S. aureus* and *S. epidermidis* biofilms lead to devastating illnesses such as prosthetic valve endocarditis.[6] Other *Staphylococcus* species have been reported to produce biofilms. *S. haemolyticus* biofilms were associated with acute conjunctivitis during treatment with punctual plugs.[7] *S. capitis* was reported in a rare case of pyomyositis in the paraspinal neck muscles of a diabetic patient.[8] *S. argenteus* strains were found in a persistent case of prosthetic joint infections caused by biofilms, host immune resistance, osteoblast invasion, and intracellular persistence.[9] *S. xylosus* have been isolated from bioaerosols in broiler chicken barns.[10] From these studies, biofilm formation is exhibited by numerous species and can cause various diseases that contribute to its medical significance.

Accumulating data suggest that staphylococcal biofilm production involves the eDNA, Polysaccharide Intercellular Adhesin (PIA), or poly-β(1,6)-N-acetyl-glucosamine (PNAG), and Microbial Surface Components Recognizing Adhesive Matrix Molecules (MSCRAMMS), which comprises the extracellular matrix component for the biofilm.[11] These microbial proteins are encoded by various biofilm-associated genes. PIA/PNAG is synthesized by the enzyme products of the intercellular adhesion (*ica*) locus, making it a good mediator for intercellular and surface adhesion and the structural integrity of biofilms *in vivo* and *in vitro*.[12] The *ica* locus encodes four proteins, IcaA, IcaD, IcaB, and IcaC, and is regulated by the Intercellular adhesion Regulatory protein (IcaR).

IcaR is a 22 kDa protein belonging to the Tetracycline (TetR) family of transcriptional regulatory proteins encoded in the *ica* locus.[13] It binds to the promoter of the *icaABCD* and 5' of the start codon of IcaA, thus regulating the expression of the *icaABCD*. In the study of Conlon *et al.* (2002), IcaR inactivation by NaCl and ethanol led to increased IcaA expression in an *S. epidermidis* isolate that produces a weak biofilm.[14] Additionally, Jefferson *et al.* (2004) assert that the deletion of the IcaR expression in *S. aureus* and *S. epidermidis* significantly increased PIA, thereby increasing bacterial adhesion and affecting biofilm

development.[15] In the findings of Yu *et al.* (2012), biofilm formation in *S. aureus* was decreased when *ica*R transcription was enhanced and activated using autoinducer-2, a Quorum-Sensing (QS) signal for interspecies communication.[16] These studies laid a strong foundation for the regulatory activity of IcaR through experimental research; however, it is notable that there is a lack of information demonstrating the exact molecular mechanisms explaining the functional characteristics of the IcaR protein in various *Staphylococcus* species.

In silico characterization and homology modeling are powerful techniques that can determine unknown protein structures.[17] According to Rabbi *et al.* (2021), *in silico* characterization uses different computational tools and servers to determine the protein sequence's physico-chemical properties, such as amino acid sequence, hydrophobicity, and overall structure and probable functional domain in query protein sequences.[18] Homology modeling uses known template structures and compares them to an undetermined homologous protein structure.[19] These techniques were utilized to investigate and evaluate the properties, structure, and function of the IcaR protein. In line with these, this study proposes that the direct relation of the phylogeny of *Staphylococcus* species indicates similarity in structure and function of IcaR protein as an *icaABCD* regulator. This study will be beneficial as it can be used as a future reference in the formulation of significant interventions that can target a specific area in the protein structure, such as stimulating the expression of genes that can inhibit the formation of biofilms.

MATERIALS AND METHODS
Protein Sequence Retrieval from NCBI (GenBank)

The IcaR crystalline structures of the *S. aureus* (3GEU) and *S. epidermidis* (2ZCN) were retrieved from the Protein Data Bank (PDB) and served as template sequences for homology modeling of query IcaR protein sequences. Available IcaR protein sequences of different *Staphylococcus* species were retrieved from the National Center for Biotechnology Information (NCBI) GenBank website (https://www.ncbi.nlm.nih.gov/). The protein sequences of *S. xylosus* (GEQ10639.1), *S. haemolyticus* (ACL68644.1), *S. argenteus* (WP_000790912.1), *S. saprophyticus* (BAE17260.1), *S. caprae* (AAK18070.1) and *S. capitis* (AHZ90663.1) were chosen for testing. Criteria for homology modeling were followed, which required a sequence identity percentage of at least 40%.[20] BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to align the query

sequences against the template sequences to determine the best match for homology modeling.[17]

Sequence Alignment and Phylogenetic Analysis

Sequence alignment was done using ClustalW in MEGA 11, which was required for homology modeling. After alignment, all IcaR protein sequences and a selected outgroup obtained from GenBank were compiled in a single FASTA file format for the creation of a phylogenetic tree. The selected outgroup was obtained via BLASTp search of IcaR protein sequences. The basis for selection involved choosing a group that did not belong to the ingroup but was relatively closely related. *S. simulans* (VED61380.1) fits this criterion and serves as a point of reference for the phylogenetic tree. The phylogenetic tree was developed using the Phylogeny section of the same application. Phylogeny of the chosen *Staphylococcus* species involved the use of bootstrap values Neighbor-Joining (NJ), Minimum Evolution (ME), and Maximum Likelihood (ML), focusing on ML with Jones-Taylor-Thornton (JTT) as the substitution model. These bootstrap values were done in 1000 replicates for increased reliability of results.

Physico-chemical Profiling

ProtParam (Primary Structures)

The chosen template and query proteins, including *S. epidermidis* (2ZCN), *S. aureus* (3GEU), *S. haemolyticus* (ACL68644.1), *S. capitis* (AHZ90663.1), *S. caprae* (AAK18070.1), *S. xylosus* (GEQ10639.1), *S. argenteus* (WP 000790912.1), and *S. saprophyticus* (BAE17260.1) were subjected to physico-chemical profiling using ProtParam (ExPASy) (https://web.expasy.org/protparam). This provided the Molecular Weight (MW), theoretical Isoelectric point (pI), Net Charge (NC), half-life in hours, Aliphatic Index (AI), Instability Index (II), and Grand Average of Hydropathy (GRAVY).[21]

SOPMA (Secondary Structures)

SOPMA (https://prabi.ibcp.fr/htm/site/web/app. php/home) was used to produce secondary structure protein structures of all the sequences.[17] The server determined the percentage of α-helices, β−sheets, and random coil structure of all protein sequences.

Tertiary and Quaternary Structure Determination

The tertiary structures of the protein sequences were constructed using SWISS-MODEL (https:// swissmodel.expasy.org/).[17] The IcaR protein sequence of *S. epidermidis* (2ZCM) and *S. aureus* (3GEU) served as the template for the creation of models for the chosen query sequences. Furthermore, i-TASSER (https://

zhanggroup.org/I-TASSER/) and Phyre2 (www.sbg. bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) were utilized to provide 3D protein structures.[22,23] The default options of the website were used. The viability of the generated structures was then verified using the WHATCHECK server of the SAVES v6.0 website (https://saves.mbi.ucla.edu/). It used the Ramachandran plot in which positions of non-Gly residues in the DA (disallowed regions) and the x-y plot of the phi/psi dihedral angles between N- and C-terminal sides of the planar peptide bonds were used to predict the possible backbone conformation of the proteins.[20] Co-regions (Most Favored Core Region) greater than 90% indicated good model quality.[24]

The chosen 3D model of the proteins was refined by the GalaxyWEB server (https://galaxy.seoklab.org) to improve the models by detecting unreliable regions and performing *ab initio* modeling. Visualization and analysis of the tertiary and quaternary protein structure were done in the Research Collaboration for Structural Bioinformatics (RCSB) 3D Protein Viewer MolStar (https://www.rcsb.org/3d-view).

RESULTS

Protein Sequence Retrieval

The IcaR protein sequences of the template and query species were retrieved from the Protein Data Bank and GenBank, respectively. Using the sequences of the biofilm-producing *Staphylococcus* species, local similarities between sequences were identified and aligned using BLASTp. IcaR protein sequences of the query species having more than 40% similarity scores with the template species were then selected for further analysis, while excluding those with less than 40% (not presented). In increasing percent similarities, *S. xylosus* (51.63), *S. saprophyticus* (52.20), *S. caprae* (76.76), *S. capitis* (77.96), and *S. haemolyticus* (98.92) had suitable protein sequences for modeling with *S. epidermidis,* while *S. argenteus* (92.47) for *S. aureus* COL.

Sequence Alignment and Phylogenetic Analysis

All template and query protein sequences were placed in one clade that comprises four lineages (Lineages I – IV), with *S. simulans* serving as the outgroup (Figure 1). Lineage I included *S. epidermidis* and *S. haemolyticus,* with all bootstrap values reaching 100. Lineage II included *S. capitis* and *S. caprae* with high bootstrap values, especially for ML. The node located between Lineages I and II also provided a decent score, notably for ME and NJ. Lineage III included *S. xylosus* and *S. saprophyticus,* and Lineage IV comprised *S. aureus* and *S. argenteus,*

both of which differ only from the ML bootstrap value. The internal node between Lineages I to III offered the lowest bootstrap values. No statistical values were produced at the root of the ingroup.

Primary Structure Physico-chemical Profiling

All amino acids were present in the species (Figure 2). Leucine (L), lysine (K), serine (S), and isoleucine (I) had the highest percentages, followed by aspartic acid (D), glutamic acid (E), asparagine (D), phenylalanine (F), and tyrosine (Y). Tryptophan (W), cysteine (C), and proline (P) had the lowest percentages in almost all species.

The physico-chemical properties (Table 1) of the IcaR protein of selected *Staphylococcus* species revealed that *S. xylosus* had the highest MW with 22,791.91 Da, while *S. saprophyticus* had the lowest MW of 20,058.75 Da. The isoelectric point of the source species ranges from 4.88 to 5.51, which means the pH is mainly acidic. All the source species displayed negatively charged residues and a half-life of 30 hr. The aliphatic index varied from 84.95 (*S. aureus* COL) to 97.91 (*S. xylosus*). The Grand Average of Hydropathy (GRAVY) of the source species ranged from -0.128 (*S. xylosus*) to -0.429 (*S. caprae*).

Secondary Structure Prediction

Figure 3 shows the percentages of the α-helices, β-sheets, and random coils of the IcaR protein of selected *Staphylococcus* species. The predicted secondary structures showed that all IcaR proteins of the *Staphylococcus* species had a high percentage of α-helices. Consequently, both the β-sheets and random coils exhibited low percentages.

Tertiary Structure Determination, Validation, and Refinement

Tertiary structures of the query protein sequences were generated using Phyre2, SWISS-MODEL, and I-TASSER and evaluated through the WHATCHECK server of the SAVES v6.0 website by inspecting the Ramachandran Plot values. As shown in Figure 4, among the modeling servers, SWISS-MODEL yielded the highest results for all the proteins. The Ramachandran model score for *S. capitis* AHZ90663.1 using the Phyre2 was 96.6%, slightly higher than the model generated from SWISS-MODEL, which yielded 96%. Both Phyre2 and SWISS-MODEL provided acceptable models that satisfied the criteria of having a value over 90%.
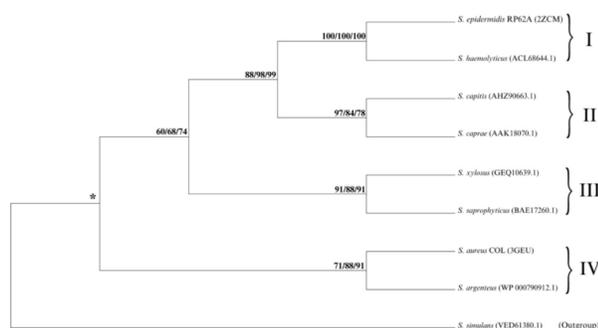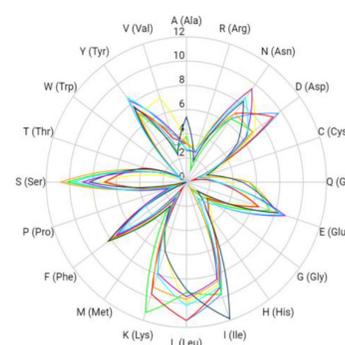

Figure 1: Phylogenetic tree based on IcaR protein of different *Staphylococcus* species with bootstrap values (ML/ME/NJ) done in 1000 replicates. Asterisks (*) denote that no statistical report was produced.

Figure 2: Amino Acid Composition per *Staphylococcus* species.

Table 1: Physico-chemical Properties of IcaR Protein of *Staphylococcus* species from ProtParam.							
IcaR Source Species	MW	AA	pI	NC	Half-life (hours)	AI	GRAVY index
S. epidermidis 2ZCM (strain RP62A)	22169.15	185	5.36	-7	30	94.86	-0.391
S. aureus 3GEU (strain COL)	21987.08	186	5.17	-4	30	84.95	-0.367
S. xylosus (GEQ10639.1)	22771.91	191	5.17	-9	30	97.91	-0.128
S. haemolyticus (ACL68644.1)	22154.14	185	5.51	-6	30	94.32	-0.393
S. argenteus (WP_000790912.1)	22111.02	186	4.95	-6	30	86.99	-0.397
S. saprophyticus (BAE17260.1)	20058.75	168	4.84	-10	30	93.51	-0.153
S. caprae (AAK18070.1)	22575.42	190	4.88	-9	30	86.16	-0.429
S. capitis (AHZ90663.1)	22056.87	186	4.95	-8	30	90.65	-0.381

Note. MW= molecular weight; AA = Number of Amino Acids; NC= net charge; AI= aliphatic index; GRAVY= grand average of hydropathicity index.
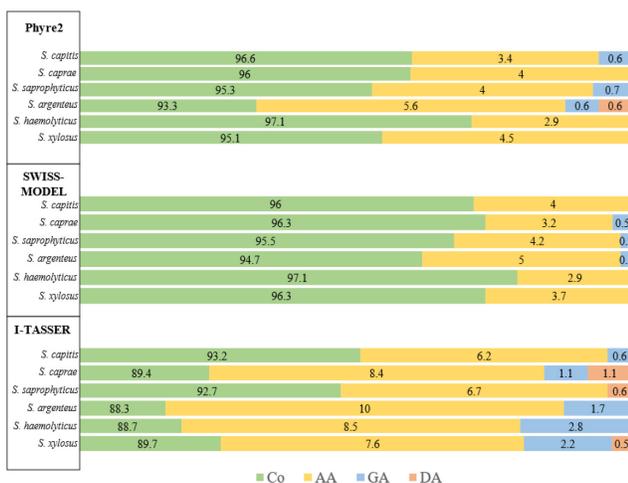
Figure 3: Secondary Structure Comparison from SOPMA.

Figure 4: Ramachandran Plot calculation using SAVES v6.0.

Note. Co= Most Favored (Core) Regions; AA= Additionally Allowed Regions; GA= Generously Allowed Regions; DA= Disallowed Regions.

Tertiary and Quaternary Structure Evaluation and Analysis

Three-dimensional protein inspections provided significant sequence homology from the N-terminal to the C-terminal of most query species when compared to either 2ZCN or 3GEU. When the A chain of 2ZCN served as the template sequence for superimposition, the Root-Mean-Squares Deviation (RMSD) for *S. caprae, S. haemolyticus, S. xylosus,* and *S. capitis* were 0.07Å, 0.08Å, 0.09Å, 0.49Å, respectively. Two species portrayed lower homology with 2ZCN as *S. argenteus* obtained a good RMSD value of 1.77Å, while *S. saprophyticus* showed bad alignment with a value of 5.13Å. In contrast, when 3GEU was used as the template for superimposition, RMSD values of *S. argenteus* and *S. saprophyticus* exhibited high homology with RMSD values of 0.07Å and 0.84Å, respectively, most likely due to their lineage proximity as shown in Figure 2.

Both the IcaR template and query models produced noticeable disordered regions from residues 65 to 75, located towards the C-terminal domain. Residues between helices α9 and α8 also contained slight variations between template groups. However, residues 127 to 137 from *S. xylosus* produced wide deviations, likely due to its higher residue amount (191). Moreover, *S. saprophyticus* lacked the helix α9, which is attributed to its lower amino acid count (168). Hence, variations in their superimposition could be due to their differing residue amount, especially *S. saprophyticus* and *S. xylosus*. Nonetheless, when query models were superimposed towards the IcaR template protein with lower RMSD values, both the N-terminal and C-terminal domains remained relatively homologous.

Quaternary comparisons mainly involved N-terminal evaluations by measuring the distance between the two Tyr38 Cα atoms located at the α3 helix of each IcaR monomer, which is indicated to bind to the adjacent major groove of DNA.[25] *S. saprophyticus* and *S. xylosus* are exceptions, as sequence alignments predict that their Tyr42 Cα atoms are more involved in this process due to their differing amino acid count. Figure 5 illustrates the narrow differences between the tyrosine molecules that bind at the major groove of DNA in each IcaR model produced, being 38.37–39.21Å. It should be noted that increased distances between dimers could inhibit *ica* operon regulation.[25]

DISCUSSION

Primary Structure Physico-chemical Profiling

Helices α8 and α9 of the TetR family form most of the dimerization interface.[26,27] Hence, the missing helix α9 observed in *S. saprophyticus* could lead to less
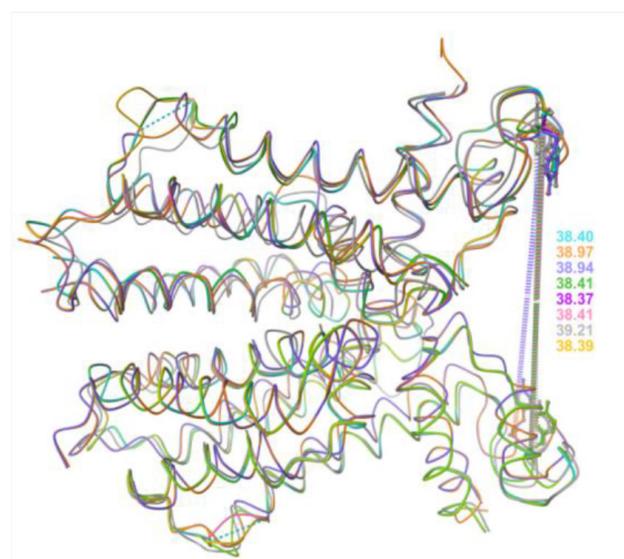

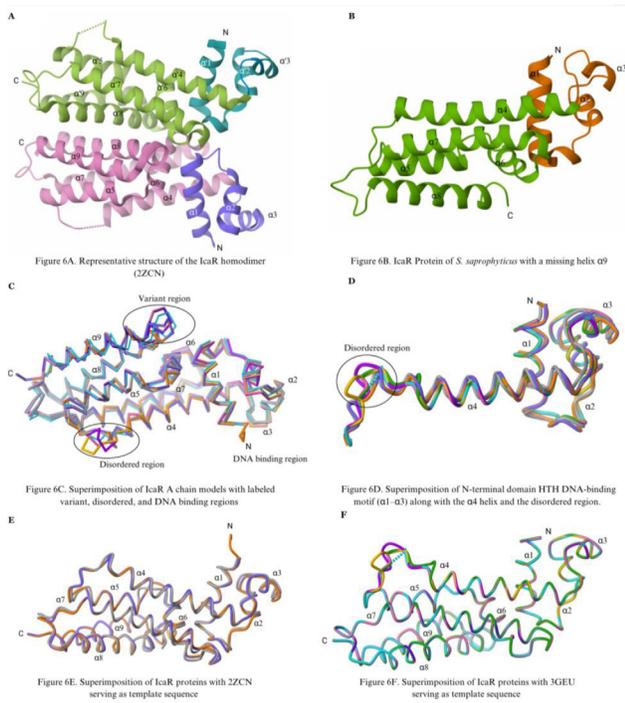Figure 5: Distance comparison of tyrosine residues in superimposed dimer structures.

Figure 6: Structure of the IcaR Protein. (6A) Representative structure of the IcaR homodimer (2ZCN). (6B) ***S. saprophyticus*** is missing the helix α9. (6C) Superimposition of IcaR A chain models. (6D) Superimposition of N-terminal domain Helix-Turn-Helix (HTH) DNA-binding motif (α1−α3) along with the α4 helix and the disordered region. (6E) Superimposition of IcaR proteins with 2ZCN serving as template sequences. (6F) Superimposition of IcaR proteins with 3GEU serving as template sequences.

stabilized dimer interaction and higher susceptibility to conformational changes that reduces its biofilm regulatory properties. This insufficient helix structure can be attributed to its low molecular weight and residue count. In contrast, the other *Staphylococcus* species retained their structures; hence, retaining their biofilm regulatory function in this aspect.

The pI and NC of IcaR proteins are influenced by the aspartate and glutamate residues (acidic side chains), resulting in low pI values and negative net charges. IcaR proteins are then more likely to reside in the cytoplasm of the bacteria that thrive in low pH and low membrane charge environments.[28] Furthermore, the more negatively charged and acidic variants of IcaR proteins would likely have decreased DNA-binding affinity due to the negatively charged DNA backbone, hence, leading to less biofilm regulation.

The *in vivo* half-life predicts the time required for half of the protein amount to be degraded.[21] Since all IcaR proteins exhibit a half-life of 30 hr, their biofilm regulation function is then suggested to be stable in the absence of external environmental factors. The proteins also show high AI, illustrating extended volumes of

aliphatic side chains that contribute to increased thermal stability.[21] Hence, these side chains may contribute to stabilized functions at increased temperatures. The low GRAVY index suggests hydrophilicity, leading to a greater affinity for hydrophilic molecules.[17] IcaR inactivation from salt and ethanol[14] may then be explained by its increased hydrophilic binding affinity to IcaR proteins.

Secondary Structure Analysis

The increased percentage of α-helices found in *S. haemolyticus* could indicate a more ordered and stable secondary structure in comparison to other IcaR proteins. In contrast, *S. capitis* is found to have more random coils in its structure, which confers flexibility but increases susceptibility to degradation since it lacks a defined structure. In addition to its purpose of inducing protein stability,[29] α-helices provide necessary functions from protein-DNA interactions via the HTH motif, dimer formation, and ligand binding.[25] Hence, increased protein folding towards more defined α-helices may indicate the maintenance of biofilm regulatory functions.

Tertiary and Quaternary Evaluation and Comparison

The C-terminal domain (α4−α9) serves as the region for dimerization, while the N-terminal (α1−α3) serves as the DNA-binding domain, especially the Lys33 and Lys34 of helix α3.[25] Helices α1−α3 create the N-terminal region of each monomer, which directly interacts with the *ica* operon through an HTH motif.[25] Despite the slightly higher sequence identity *S. saprophyticus* (BAE17260.1) produced towards 2ZCN, the lower RMSD value calculated and lineage proximity indicates that 3GEU serves as a closer homolog. Thus, results suggest that the IcaR proteins of *S. aureus, S. saprophyticus, and S. argenteus* create similar biofilm regulatory functions. In contrast, *S. epidermidis, S. caprae, S. haemolyticus, S. xylosus,* and *S. capitis* are exceptionally homologous, suggesting similar *ica* operon regulation. These similarities in structure align with the study's phylogenetic analysis that illustrates lineage proximity has led to higher homology between IcaR proteins. The C-terminal region of IcaR proteins contains variations in the structure, leading to slight differences in dimerization and ligand binding. The small, disordered region shown in the C-terminal domain (Figure 6D) likely provides various allowable conformations, which suggests flexibility that is ideal for efficient dimer formation.

Given the narrow range of 38.37−38.41Å between the two tyrosine Cα atoms in all IcaR dimers (Figure 5),

the arrangement of HTH motifs is conserved. This N-terminal homology possibly denotes preserved DNA binding functions between species, especially among closer lineages. There are also 13 other amino acid residues involved in DNA binding, which include M1, K2, T22, L23, I32, K33, K34, A35, S36, Y38, Y39, H40, N43, and K44.[25] Homology has remained throughout most IcaR proteins, except for residue 43 (asparagine) being replaced with serine in *S. aureus* and *S. argenteus*. The effects of antibiotics on IcaR have been established, indicating that streptomycin and gentamicin increased the distance between the two HTH motifs, which inhibited its binding capability.[25] Thus, by interfering with IcaR binding to DNA, biofilm formation may be elicited, particularly for *S. epidermidis*. The high homology of IcaR proteins towards 2ZCN (*S. epidermidis*) produced in this study suggests that gentamicin and streptomycin may also induce similar effects on the HTH motifs. Nevertheless, further experimental research is required to confirm these findings.

CONCLUSION

In this study, the physico-chemical properties, molecular homology, and functional implications of the IcaR protein structures across relevant biofilm-producing *Staphylococcus* species have been elucidated. IcaR protein activity remains relatively conserved with varying differences. In the N-terminal DNA binding region, there was a narrow difference in distance between the tyrosine residues that bind to the DNA across all IcaR proteins, indicating slight variations in function. The variation in residue type towards necessary amino acids in the region was minor, as both *S. aureus* and *S. argenteus* only exhibited one residue disparity from other IcaR proteins. These findings support the idea that the N-terminal region has a conserved *ica* operon regulatory function. In contrast, the C-terminal region introduced high variations in various areas between other α-helices (e.g., disordered regions). The difference in residue type in these portions suggests a disparity of function towards dimerization and ligand binding. Ultimately, these implications suggest that certain inhibitory or stimulatory ligands could have different affinities toward specific C-terminal regions dependent on species type.

Homology modeling has allowed the creation of unknown IcaR proteins into reliable models with newly determined structural and functional properties. The use of presently available web servers and software has allowed for the molecular determination of proteins. Three-dimensional analysis of proteins supported findings in phylogenetic relationships between IcaR proteins. This data can then be used to develop novel strategies for regulating IcaR expression and provide reliable 3D models as biochemical references that will guide future research. Nonetheless, *in silico* homology modeling provided the foundation for nuanced molecular analysis and can pave the way for future experimental studies to compare theoretical and empirical data.

RECOMMENDATION

Based on the findings of this study, the following recommendations are suggested. First, experimental research (e.g., crystallography) can confirm the study's findings by determining the structural and functional capabilities of each IcaR protein from the selected *Staphylococcus* species. Second, the exploration of therapeutic strategies that specifically target C-terminal regions of different IcaR proteins can determine the pharmacologic effects of the different *Staphylococcus* species. Lastly, future researchers can investigate the interactions between IcaR proteins and other regulatory proteins that are involved in PIA regulation.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ABBREVIATIONS

AA: Additionally Allowed Regions; **AI:** Aliphatic Index; **Co:** Core Regions; **EPS:** Extracellular Polymeric Substances; **DA:** Disallowed Regions; **DRI:** Device-Related Infections; **GA:** Genereously Allowed Regions; **GRAVY:** Grand Average of Hydropathy; **HAI:** Hospital-Acquired Infections; **HTH:** Helix-Turn-Helix; **ica:** Intercellular Adhesion*;* **IcaR:** Intercellular

Adhesion Regulatory Protein; **II:** Instability Index; **JTT:** Jones-Taylor-Thorton; **ME:** Minimum Evolution; **ML:** Maximum Likelihood; **MW:** Molecular Weight; **NC:** Net Charge; **NCBI:** National Center for Biotechnology Information; **NJ:** Neighbor-Joining; **PDB:** Protein Data Bank; **pI:** Isoelectric Point; **PIA:** Polysaccharide Intercellular Adhesin; **PNAG:** Poly-β(1,6)-N-acetyl-glucosamine; **MSCRAMMS:** Microbial Surface Components Recognizing Adhesive Matrix Molecules; **RMSD:** Root-Mean-Squares Deviation.

SUMMARY

Biofilm formation among *Staphylococcus* species is regulated by various biofilm-associated genes, notable of which is the intercellular adhesion (*ica*) locus which synthesizes the microbial protein PIA responsible for the structural integrity and surface adhesion of biofilms. The *ica* locus is negatively regulated by the intercellular adhesion regulatory protein (IcaR) by binding to the promoter region of the *icaABCD* and 5' of the start codon of IcaA. Despite the notoriety of biofilm-producing *Stapylococcus* species in causing HAIs, information is lacking towards the IcaR proteins of other Staphylococci e.g., *S. xylosus, S. haemolyticus, S. argenteus, S. saprophyticus, S. caprae,* and *S. capitis*. The retrieved IcaR protein sequences from the NCBI GenBank were subjected to physico-chemical profiling using ProtParam (ExPASy), secondary structure determination through SOPMA, and homology modeling using SWISS-MODEL, i-TASSER, and Phyre2. The best models determined via Ramachandran Plot analysis were refined through the GalaxyWEB server. The models were visualized and analyzed with RCSB Molstar. The IcaR proteins of *S. aureus, S. saprophyticus,* and *S. argenteus* displayed similar biofilm regulatory functions. In contrast, *S. epidermidis, S. caprae, S. haemolyticus, S. xylosus,* and *S. capitis* are exceptionally homologous, suggesting similar *ica* operon regulation. The conserved structures were proportional to their lineage proximity in the Phylogenetic tree. The N-terminal had retained most of its homology while the C-terminal region has produced the greatest variations between the selected *Staphylococcus* species, indicating variation in dimer stability and ligand susceptibility. The theoretical data generated can contribute to understanding existing biofilm regulation literature and in developing interventions for biofilm-related nosocomial infections.

AUTHORS' CONTRIBUTIONS

REFERENCES

1. Flemming HC, Wingender J, Szewzyk U, Steinberg P, Rice SA, Kjelleberg S. Biofilms: an emergent form of bacterial life. Nat Rev Microbiol. 2016;14(9):563-75. doi: 10.1038/nrmicro.2016.94, PMID 27510863.

2. Costerton JW, Lewandowski Z, Caldwell DE, Korber DR, Lappin-Scott HM. Microbial biofilms. Annu Rev Microbiol. 1995;49:711-45. doi: 10.1146/annurev.mi.49.100195.003431, PMID 8561477.

3. Assefa M, Amare A. Biofilm-associated multi-drug resistance in hospital-acquired infections: a review. Infect Drug Resist. 2022;15:5061-8. doi: 10.2147/IDR.S379502, PMID 36068834.

4. Percival SL, Suleman L, Vuotto C, Donelli G. Healthcare-associated infections, medical devices and biofilms: risk, tolerance and control. J Med Microbiol. 2015;64(4):323-34. doi: 10.1099/jmm.0.000032, PMID 25670813.

5. Moormeier DE, Bayles KW. *Staphylococcus aureus* biofilm: a complex developmental organism. Mol Microbiol. 2017;104(3):365-76. doi: 10.1111/mmi.13634, PMID 28142193.

6. Rupp ME. Clinical characteristics of infections in humans due to *Staphylococcus epidermidis*. Methods Mol Biol. 2014;1106:1-16. doi: 10.1007/978-1-62703-736-5_1, PMID 24222451.

7. Yokoi N, Okada K, Sugita J, Kinoshita S. Acute conjunctivitis associated with biofilm formation on a punctal plug. Jpn J Ophthalmol. 2000;44(5):559-60. doi: 10.1016/s0021-5155(00)00214-8, PMID 11033137.

8. Young N, Bhally H. Bilateral neck pyomyositis caused by *Staphylococcus capitis* and *Staphylococcus saccharolyticus* in a diabetic adult. Case Rep Infect Dis. 2017;2017:3713212. doi: 10.1155/2017/3713212, PMID 29109878.

9. Diot A, Dyon-Tafani V, Bergot M, Tasse J, Martins-Simões P, Josse J, *et al*. Investigation of a *Staphylococcus argenteus* Strain involved in a chronic prosthetic-joint infection. Int J Mol Sci. 2020;21(17):6245. doi: 10.3390/ijms21176245, PMID 32872360.

10. Vela J, Hildebrandt K, Metcalfe A, Rempel H, Bittman S, Topp E, *et al*. Characterization of *Staphylococcus xylosus* isolated from broiler chicken barn bioaerosol. Poult Sci. 2012;91(12):3003-12. doi: 10.3382/ps.2012-02302, PMID 23155006.

11. Chen Q, Xie S, Lou X, Cheng S, Liu X, Zheng W, *et al*. Biofilm formation and prevalence of adhesion genes among *Staphylococcus aureus* isolates from different food sources. MicrobiologyOpen. 2020;9(1):e00946. doi: 10.1002/mbo3.946, PMID 31769202.

12. Lister JL, Horswill AR. *Staphylococcus aureus* biofilms: recent developments in biofilm dispersal. Front Cell Infect Microbiol. 2014;4:178. doi: 10.3389/fcimb.2014.00178, PMID 25566513.

13. Cue D, Lei MG, Luong TT, Kuechenmeister L, Dunman PM, O'Donnell S, *et al*. Rbf promotes biofilm formation by *Staphylococcus aureus* via repression of icaR, a negative regulator of icaADBC. J Bacteriol. 2009;191(20):6363-73. doi: 10.1128/JB.00913-09, PMID 19684134.

14. Conlon KM, Humphreys H, O'Gara JP. icaR encodes a transcriptional repressor involved in environmental regulation of ica operon expression and biofilm formation in *Staphylococcus epidermidis*. J Bacteriol. 2002; 184(16):4400-8. doi: 10.1128/JB.184.16.4400-4408.2002, PMID 12142410.

15. Jefferson KK, Pier DB, Goldmann DA, Pier GB. The Teicoplanin-associated locus Regulator (TcaR) and the Intercellular adhesin locus Regulator (IcaR) are transcriptional inhibitors of the ica locus in *Staphylococcus aureus*. J Bacteriol. 2004;186(8):2449-56. doi: 10.1128/JB.186.8.2449-2456.2004, PMID 15060048.

16. Yu D, Zhao L, Xue T, Sun B. *Staphylococcus aureus* autoinducer-2 quorum sensing decreases biofilm formation in an icaR-dependent manner. BMC Microbiol. 2012;12:288. doi: 10.1186/1471-2180-12-288, PMID 23216979.

17. Alaidarous M. *In silico* structural homology odelling and characterization of multiple N-terminal domains of selected bacterial Tcps. PeerJ. 2020;8:e10143. doi: 10.7717/peerj.10143, PMID 33194392.

18. Rabbi MF, Akter SA, Hasan MJ, Amin A. *In silico* characterization of a hypothetical protein from Shigella dysenteriae ATCC 12039 reveals a pathogenesis-related protein of the Type-VI secretion system. Bioinform Biol Insights. 2021;15:11779322211011140. doi: 10.1177/11779322211011140, PMID 33994781.

19. Gromiha MM, Nagarajan R, Selvaraj S. Protein structural bioinformatics: an overview. Encyclopedia of Bioinformatics and Computational Biology; 2019:445-59.

20. Al-Khayyat MZ, Al-Dabbagh AG. *In silico* Prediction and Docking of tertiary Structure of LuxI, an Inducer synthase of *Vibrio fischeri*. Rep Biochem Mol Biol. 2016;4(2):66-75. PMID 27536699.

21. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al*. Protein identification and analysis tools on the Expasy server Walker JM editor. The proteomics protocols handbook. Springer Protoc Handb. 2005:571-607. doi: 10.1385/1-59259-890-0:571.

22. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5(4):725-38. doi: 10.1038/nprot.2010.5, PMID 20360767.

23. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845-58. doi: 10.1038/nprot.2015.053, PMID 25950237.

24. Singh R, Gurao A, Rajesh C, Mishra SK, Rani S, Behl A, *et al*. Comparative modeling and mutual docking of structurally uncharacterized heat shock protein 70 and heat shock factor-1 proteins in water buffalo. Vet World. 2019;12(12):2036-45. doi: 10.14202/vetworld.2019.2036-2045, PMID 32095057.

25. Jeng WY, Ko TP, Liu CI, Guo RT, Liu CL, Shr HL, *et al*., Shr HL, Wang AH. Crystal structure of IcaR, a repressor of the TetR family implicated in biofilm formation in *Staphylococcus epidermidis*. Nucleic Acids Res.. 2008;36(5):1567-77. Doi: 10.1093/nar/gkm1176., PMID: 18208836.

26. Routh MD, Su CC, Zhang Q, Yu EW. Structures of AcrR and CmeR: insight into the mechanisms of transcriptional repression and multi-drug recognition in the TetR family of regulators. Biochim Biophys Acta. 2009;1794(5):844-51. doi: 10.1016/j.bbapap.2008.12.001, PMID 19130905.

27. Kang SM, Kim DH, Jin C, Ahn HC, Lee BJ. The crystal structure of AcrR from *Mycobacterium tuberculosis* reveals a one-component transcriptional regulation mechanism. FEBS Open Bio. 2019;9(10):1713-25. doi: 10.1002/2211-5463.12710, PMID 31369208.

28. Tokmakov AA, Kurotani A, Sato KI. Protein pI and intracellular localization. Front Mol Biosci. 2021;8:775736. doi: 10.3389/fmolb.2021.775736, PMID 34912847. PMCID PMC8667598.

29. Li Y, Liu X, Zhu Y, Zhou X, Cao C, Hu X, *et al*. Bioinformatic prediction of epitopes in the Emy162 antigen of *Echinococcus multilocularis*. Exp Ther Med. 2013;6(2):335-40. doi: 10.3892/etm.2013.1142, PMID 24137185.